Tinghui Zhu

darthzhu@gmail.com | github.com/DarthZhu

Education

Fudan University 2022/09 – 2025/06
Master of Science in Computer Science Shanghai, China
- Advisor: Yanghua Xiao

Fudan University 2018/09 – 2022/06
Bachelor of Science in Computer Science Shanghai, China
- Undergraduate Researcher @ Knowledge Works Research Laboratory

Research Papers

Under Review

1. *Towards Visual Taxonomy Expansion*
   Tinghui Zhu, Jingping Liu, Haiyun Jiang, Yanghua Xiao, Zongyu Wang, Rui Xie, Yunsen Xian
2. *SLR:A Million-Scale Comprehensive Crossword Dataset for Simultaneous Learning and Reasoning*
   Chao Wang, **Tinghui Zhu**, Jingping Liu, Yanghua Xiao

Publications

1. *End-to-end Entity Linking with Hierarchical Reinforcement Learning* [AAAI 23]
   Lihan Chen, **Tinghui Zhu**, Jingping Liu, Jiaqing Liang, Yanghua Xiao

Work Experience

Meituan 2022/06 – 2023/03
NLP Researcher Intern, NLU team Shanghai, China
- **Hypernymy Detection & Taxonomy Correction.** Extracted hypernymy relations from query-click logs and corrected false relations in existing taxonomy. Utilized product images to improve the performance of hypernymy detection and designed hypernymy detection method that allows continuous growth of extracted hypernymy.
- **Taxonomy Construction.** Constructed a visual domain taxonomy from 40k given terms with high accuracy. Designed an unsupervised pipeline for domain taxonomy construction, significantly improving the efficiency and performance of taxonomy construction.

Projects

ChineseBlip2 2023/04 – Now
- Finetuned BLOOMZ (7b) using LORA on multilingual corpus dataset. Instruction-finetuned BLOOMZ+LORA to unlock the ability to perform unseen tasks. Finetuned the instruction-finetuned model to perform multi-rounds chat.
- Translated multiple image-text datasets using ChatGPT to enhance the generalizability of existing Chinese datasets.
- Pretrained Blip2 based on the instruction-finetuned BLOOMZ+LORA. During stage 1, aligned image-text pairs on translated and existing Chinese datasets. During stage 2, constructed a instruction finetuning dataset and designed new input formatss for instruction-finetune LLM.

Crosswords QA 2022/01 – 2022/06
- Performed research on improving performance of clue understanding based on SLR dataset.
- where common knowledge and interpretation are required.
- The goal was to challenge comprehensive crosswords filling and multiple models and learning methods were tested and adapted.

Author Disambiguation May 2020 - Jun 2020
- Applied multiple methods to distinguish authors who may have the same or wrong names.
- The project was to automatically tell them apart.
- Graph algorithms based on co-author graph, regularized expression and clustering were used to improve the performance.